



Recognizing Behaviors in Physical Space using Imager-Enabled Sensor Networks

Andreas Savvides
Embedded Networks and Application
Lab(ENALAB)
Yale University

October 26, 2005

Research Supported by:



Oki Semiconductor

My View

It's about using technology to build a bidirectional interface between the physical and cognitive world !!!

Want to selectively see, listen and react

Priority: Leverage technology and start parsing the physical world !!!

Its not about collecting and moving data around, sensor networks are there to tell a story, sometime in real time.

SAGA: Sensors As Games Architecture

The physical world is becoming the game!

**Challenge: Build a large asynchronous
computer to play the game of everyday
life...**

Many Applications Require Understanding Behaviors in Physical Space

- **Workplace Safety**

- In a lab, it could enforce safety conditions on ventilation, use of chemicals etc
- At an industrial site, a sensor network can ensure that workers don't walk under a crane, or in an empty elevator shaft
- At a childcare center, sensors could ensure that children don't walk away or that they do not get in a fight

- **Security**

- Make proactive, autonomous security systems that do better filtering of sensed data

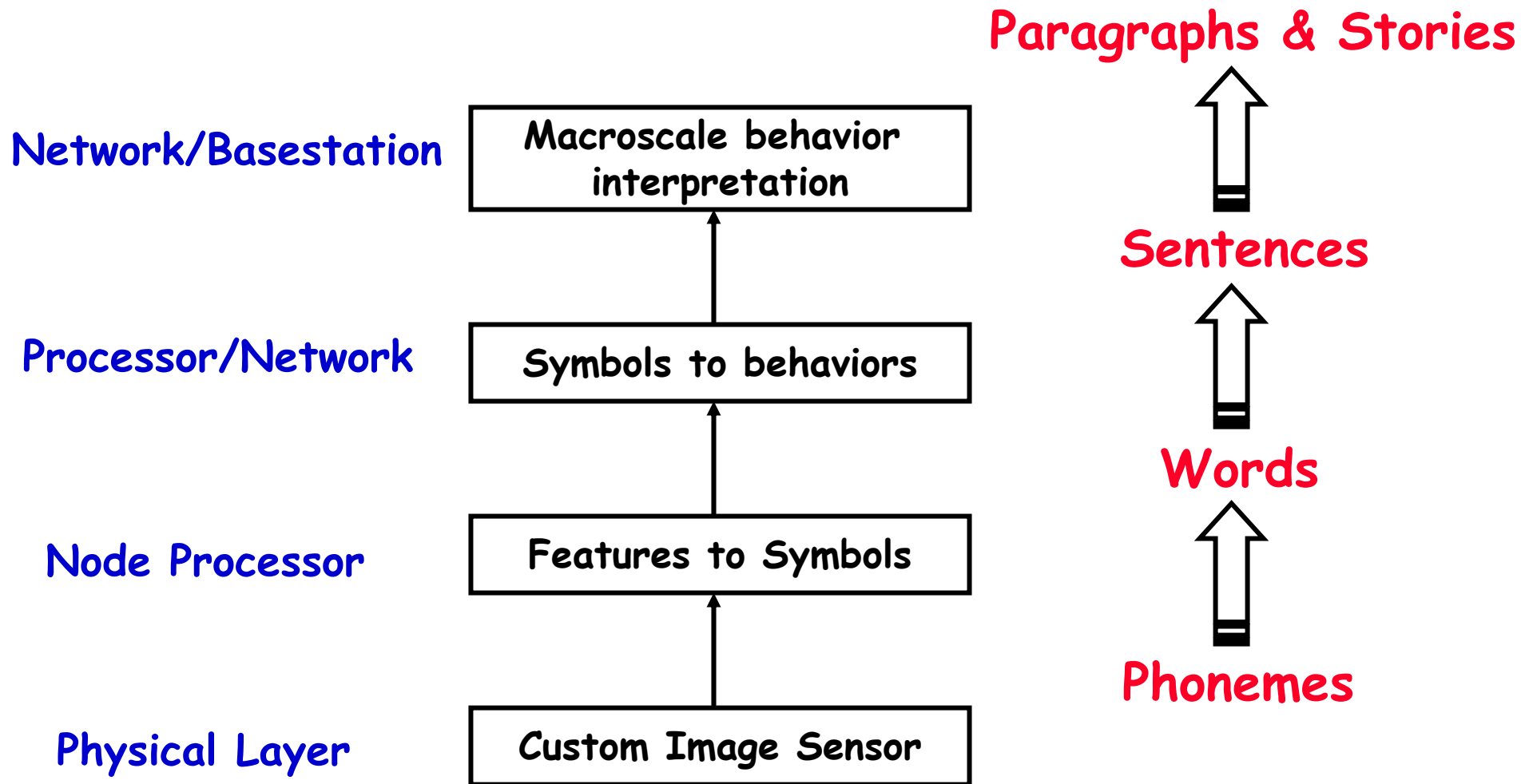
- **Entertainment**

- Sensors would be used to play games, exercise plans or education

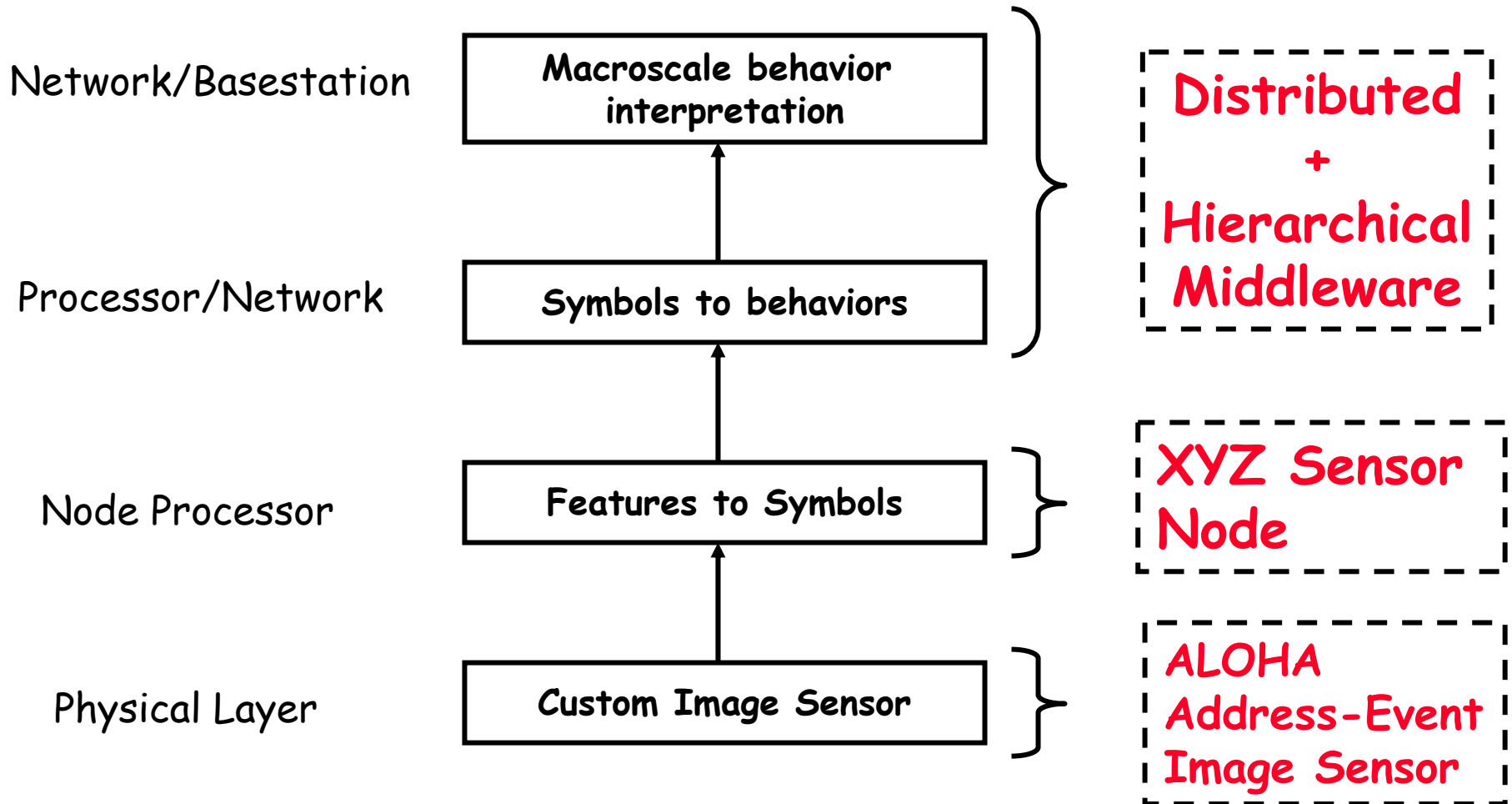
Creating a network that understands behaviors

- Use small cameras as information rich sensing modalities
 - Motions, directions, locations, events
- **Parse activities** in physical space as they happen
- Operate over low bandwidth wireless links at low latency
- Conserve power => minimum bit behavior interpretation **starting at the sensor!**

Minimum Bit Behavior Interpretation Framework

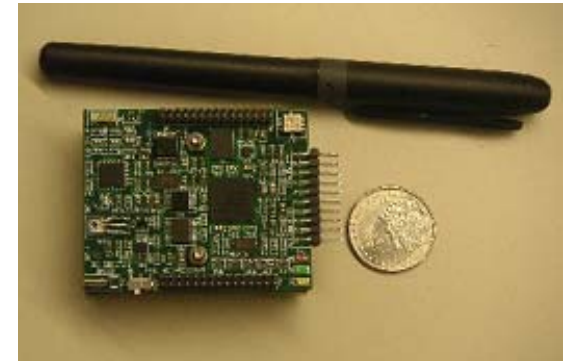


Minimum Bit Behavior Interpretation Framework

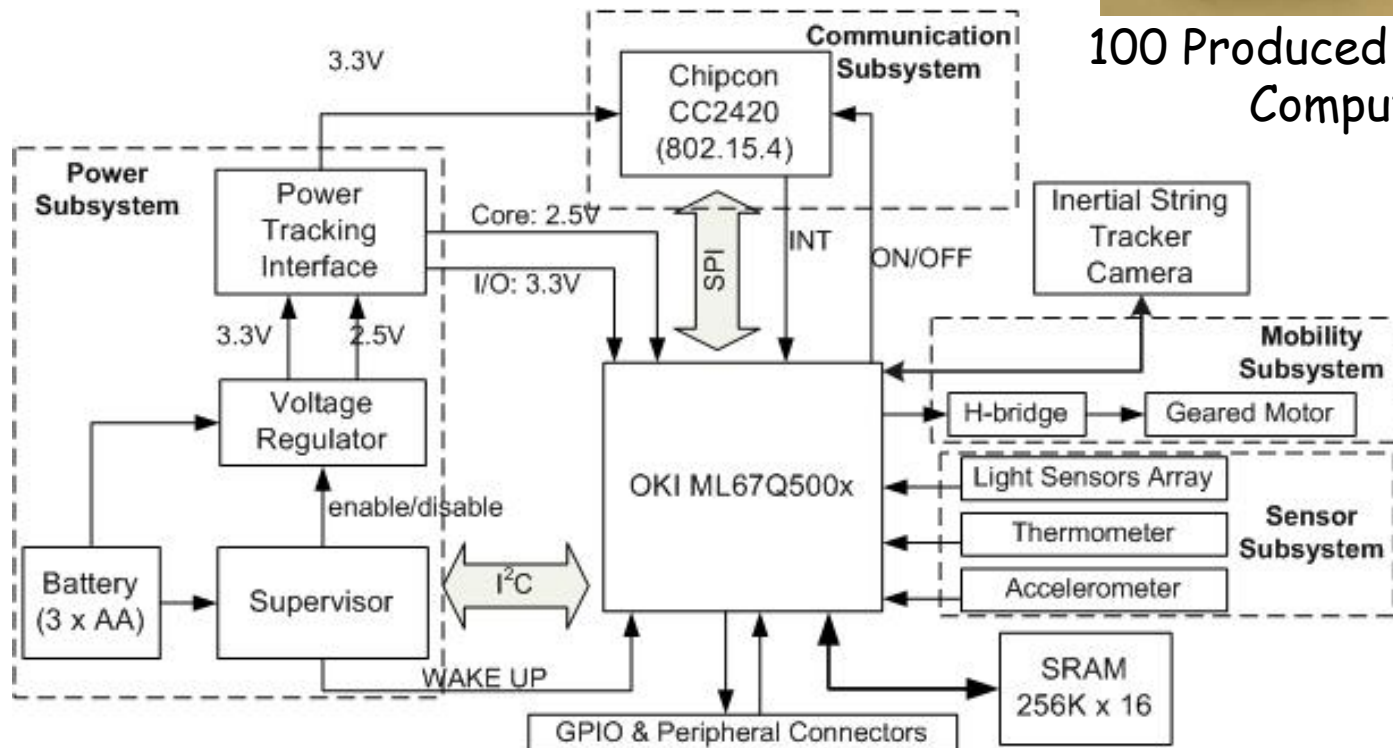


The XYZ Sensor Node Architecture

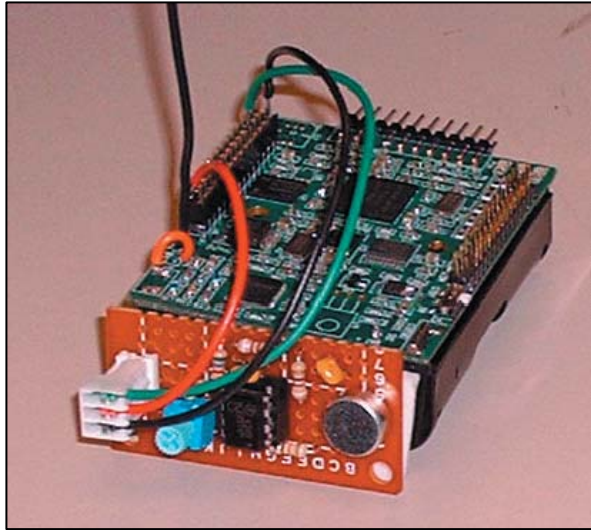
- **Chipcon CC2420, IEEE 802.15.4**
- **OKI ARM Thumb Processor**
 - 256KB FLASH, 32KB RAM, 128K x 16 External RAM
 - Max clock speed 58MHz, scales down to 1.8 MHz
 - Multiple power management functions
- **Accelerometer, temperature and light sensors**
- **Low-power sleep modes**



100 Produced by Cogent Computer

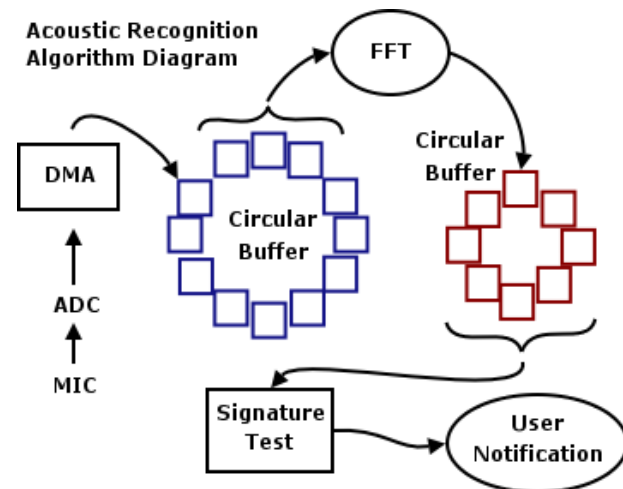
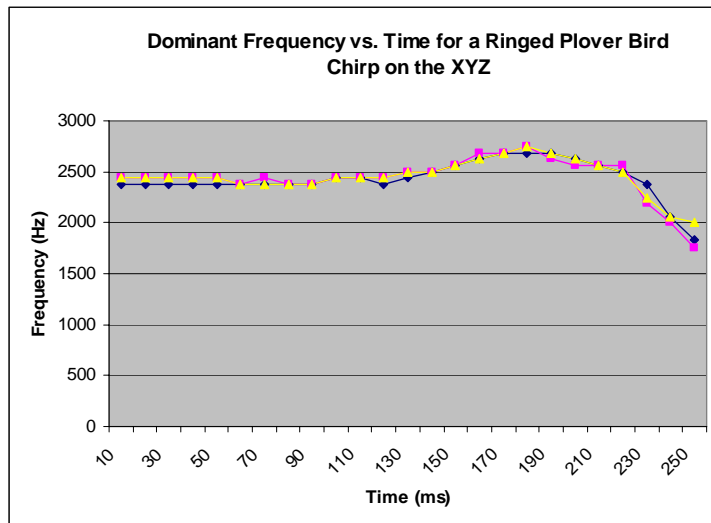


Acoustic Signature Detection on XYZ



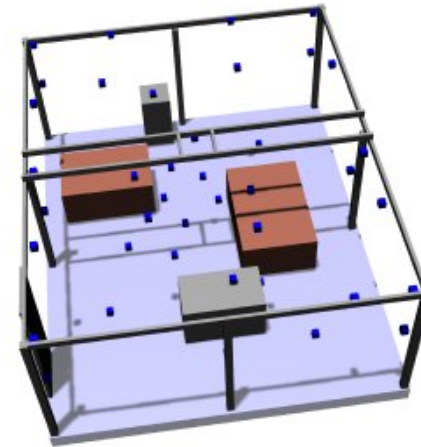
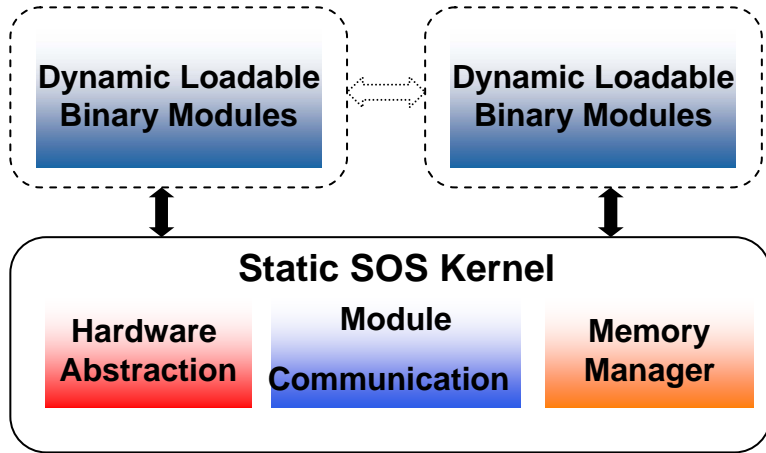
□ Prototype status

- Can recognize specific sound signatures
- Continuous sampling and processing of acoustic events up to 40KHz
- Uses a 512-Point FFT that runs in $O(1.8ms)$ on XYZ



ENALAB's Testbed

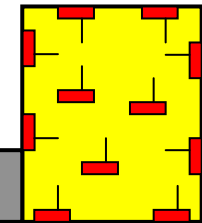
SOS Operating System
(<http://nesl.ee.ucla.edu/sos>)



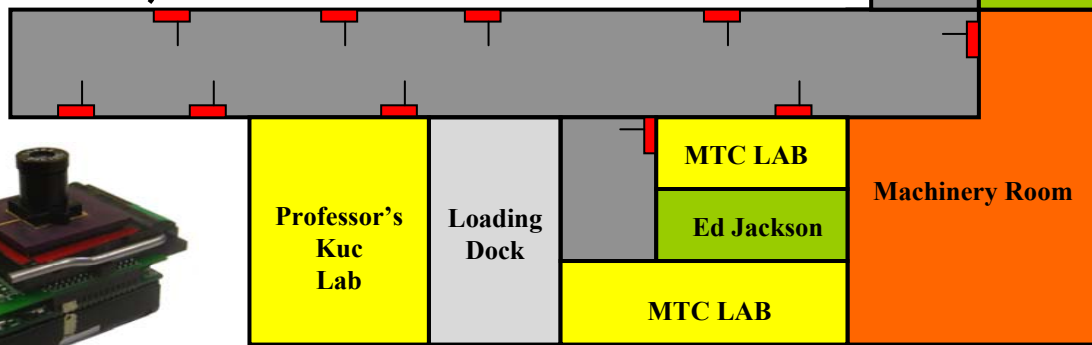
50 XYZ Nodes

Becton Center

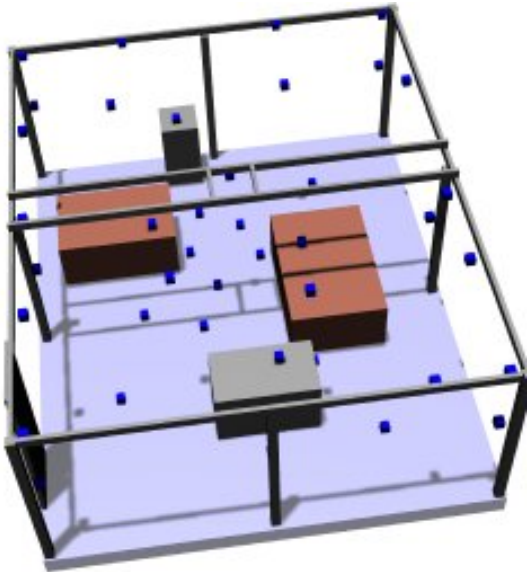
To Davies Auditorium



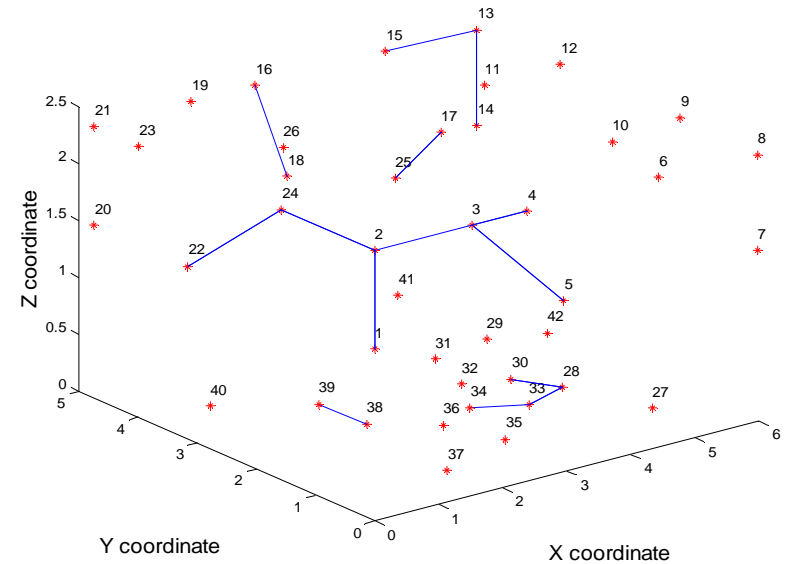
AKW #000
ENALAB



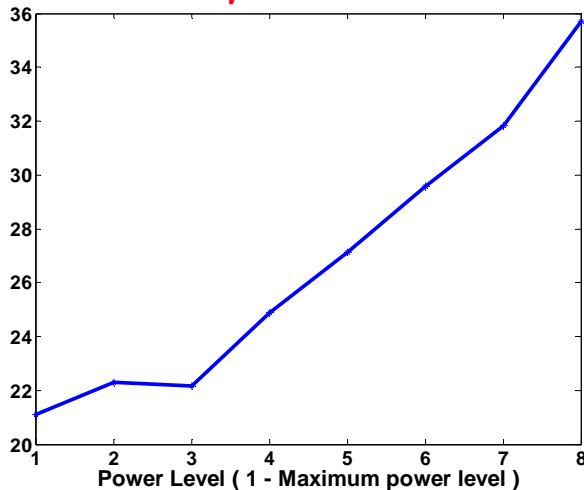
RSSI Characterization Study



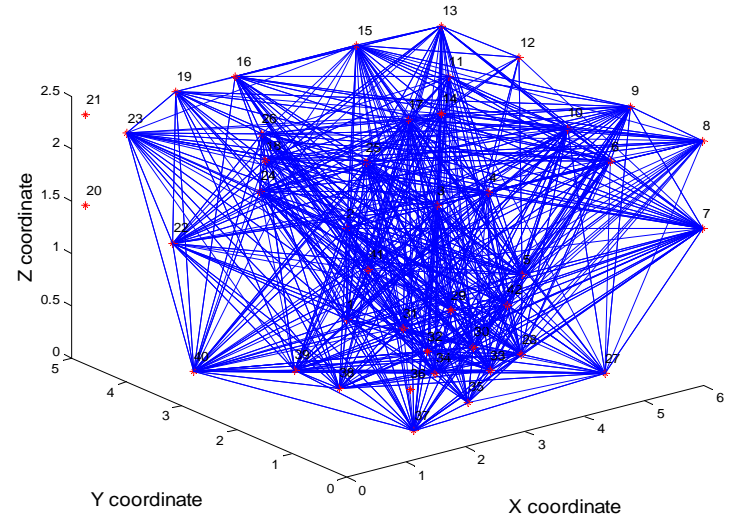
Connectivity at Power Level 7



Reconfiguration Success: 150,000 measurements - approx. 2 hours
Connectivity & RSSI Seriously Impacted by Antenna Orientation!!!



Connectivity at Power Level 4



XYZ and OV COTS Module

VGA (640x480), QVGA(320x240)

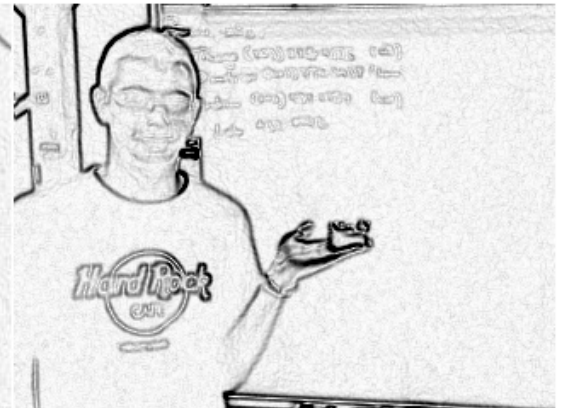
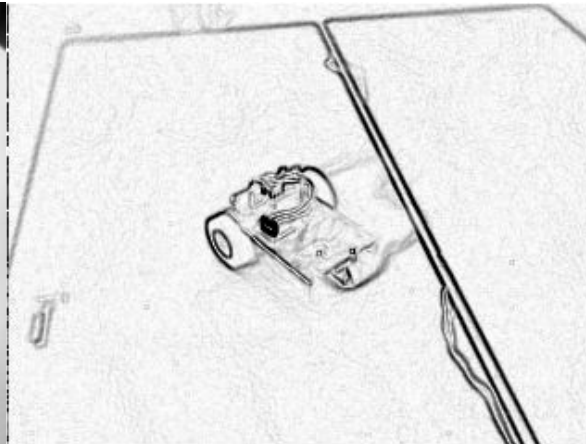
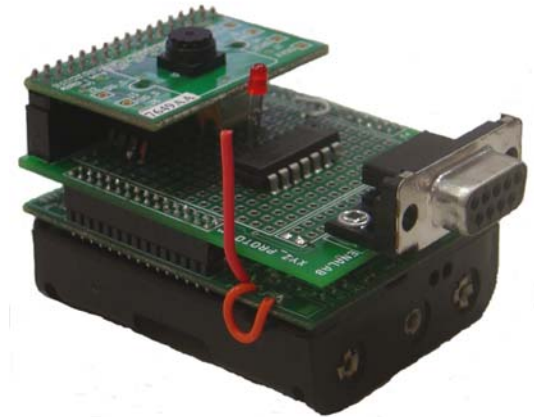
Color: 16-bits/pixel (YUV,RGB)

Sobel Edge Detection: 1bit/pixel < 500ms

Max FPS on XYZ: 4

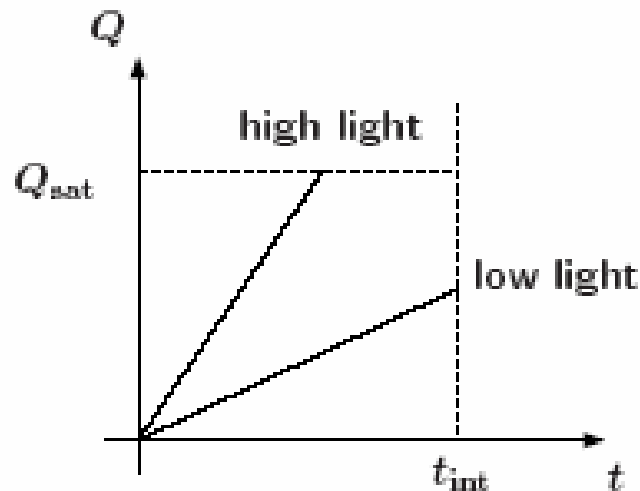
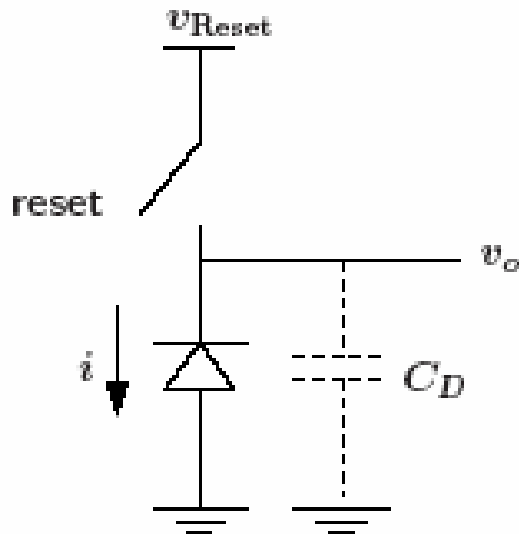
XYZ can store 1 color QVGA & 11 sobel frames

Possible to select a smaller window



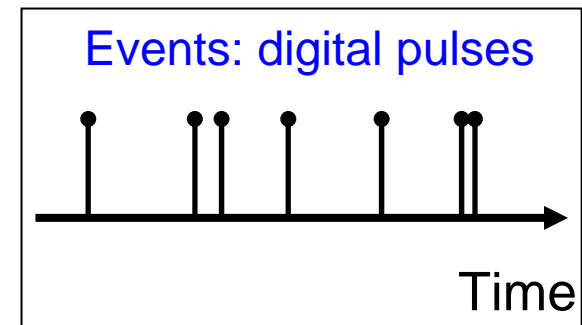
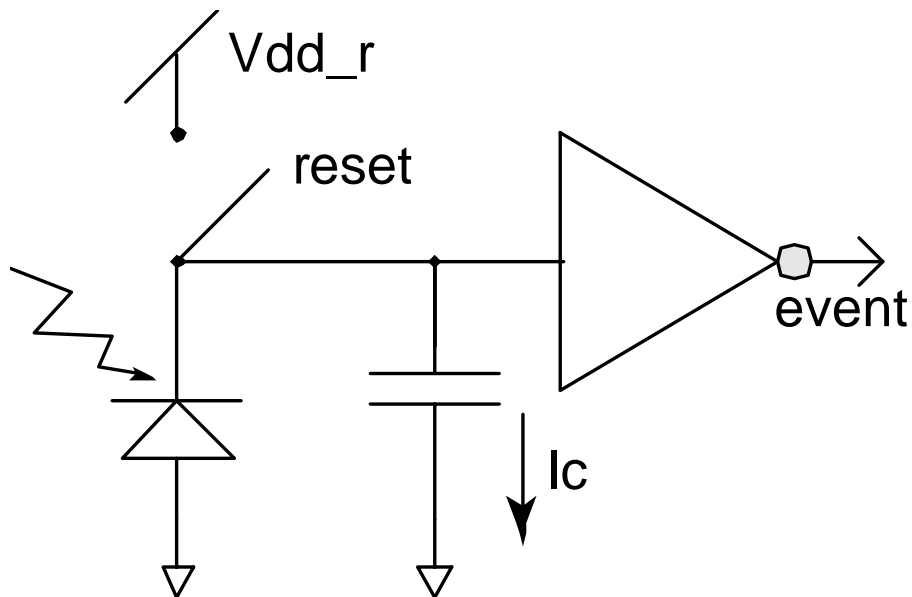
Imagers Conventional Image Sensors

- Integrate light on a capacitor for a fixed time
- Sample the **analog** capacitor voltage
- Pixels are **synchronously** scanned



Time-Domain Image Sensors

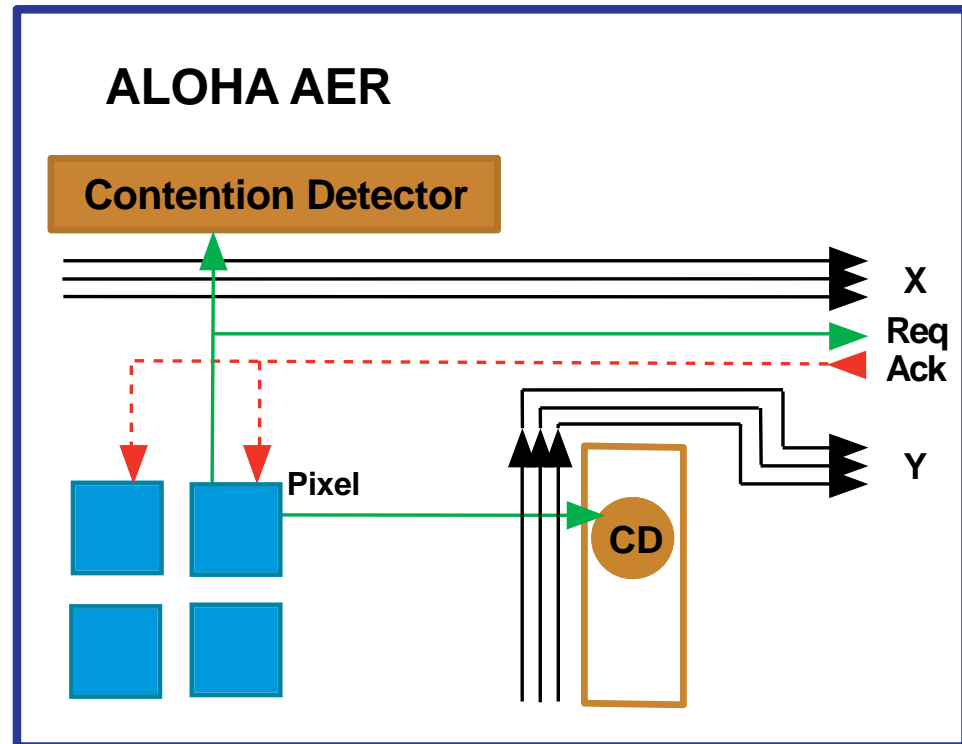
- Measure the time to integrate to a fixed voltage
- Light triggers a **digital** event
- Integrate (to threshold) and fire



Event Driven!

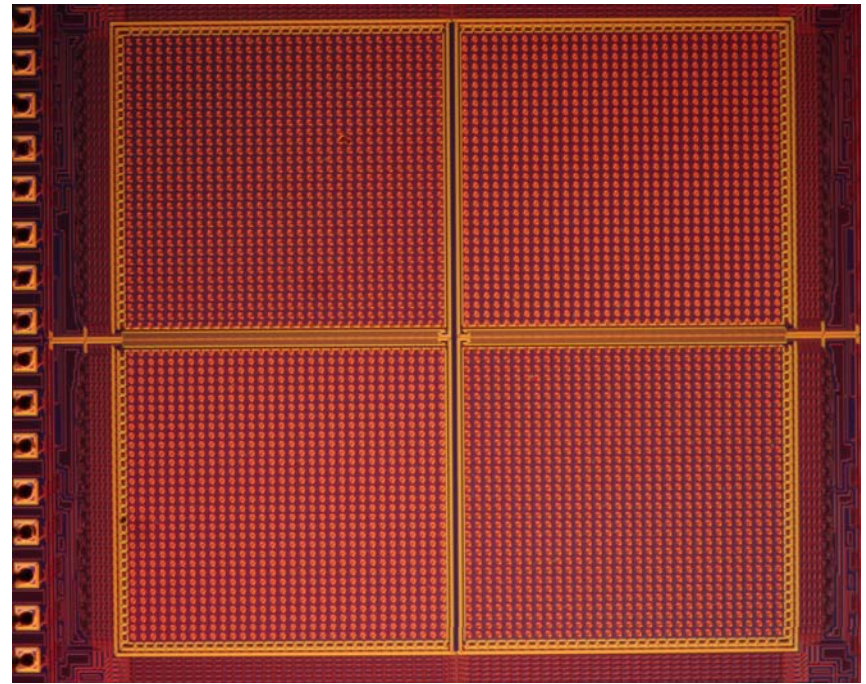
ALOHA Image Sensor

- High dynamic range
- Low power operation
- Pixel data: 1b/event
- Digital output
- No ext. components
- Asynchronous circuits
- ALOHA access
- Detects bus collisions



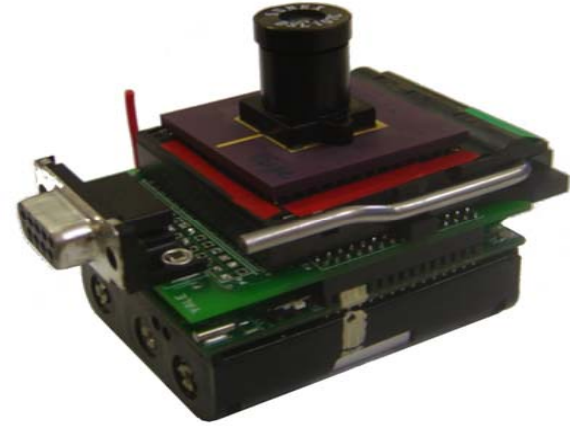
ALOHA 4x imager Performance

Technology	0.6 μ m 3M CMOS
Array Size	32 x 32, 4 quadrants
Pixel Size	32.7 μ m x 29.7 μ m
Fill Factor	6.5%
Sensor Core Size	1.2 x 1.2mm/quadrant
Dynamic Range	235dB (Pixel) 175dB (Array)
Bandwidth	8.13 μ Hz - 5MHz (Pixel) 8.33mHz - 5MHz (Array)
Sensitivity [Hz/W/m ²]	5.6x10 ⁵ (Array) 9.1 x10 ¹¹ (Pixel)
Max. FPS	2.44K (effective)
Digital Power	5.7 μ W at 2.35V, 3fps
Analog Power	7.2nW at 2.35V, 3fps

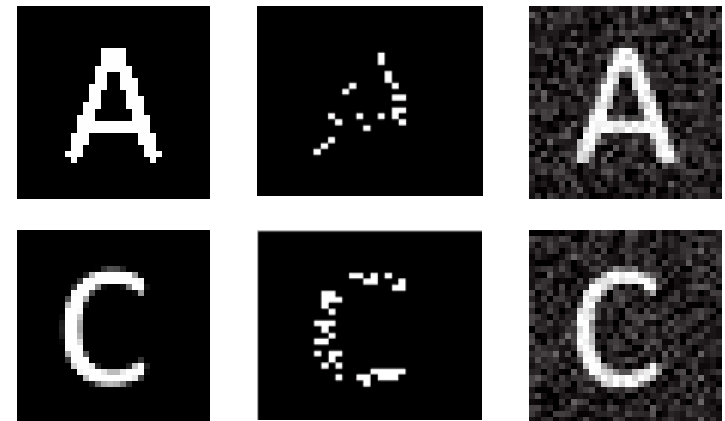
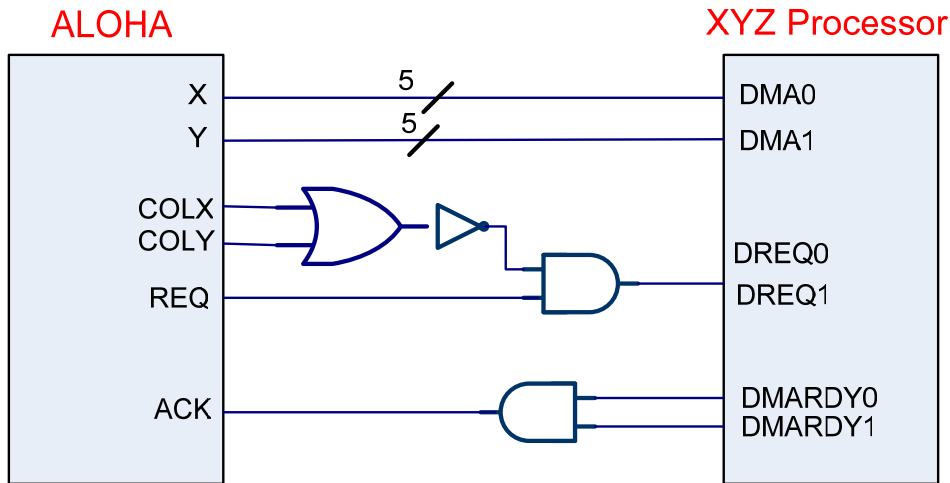


XYZ + ALOHA

- Image formation around 500 events
- Event-driven processing
- Good energy latency & selectivity



ALOHA + XYZ < 20ms detection

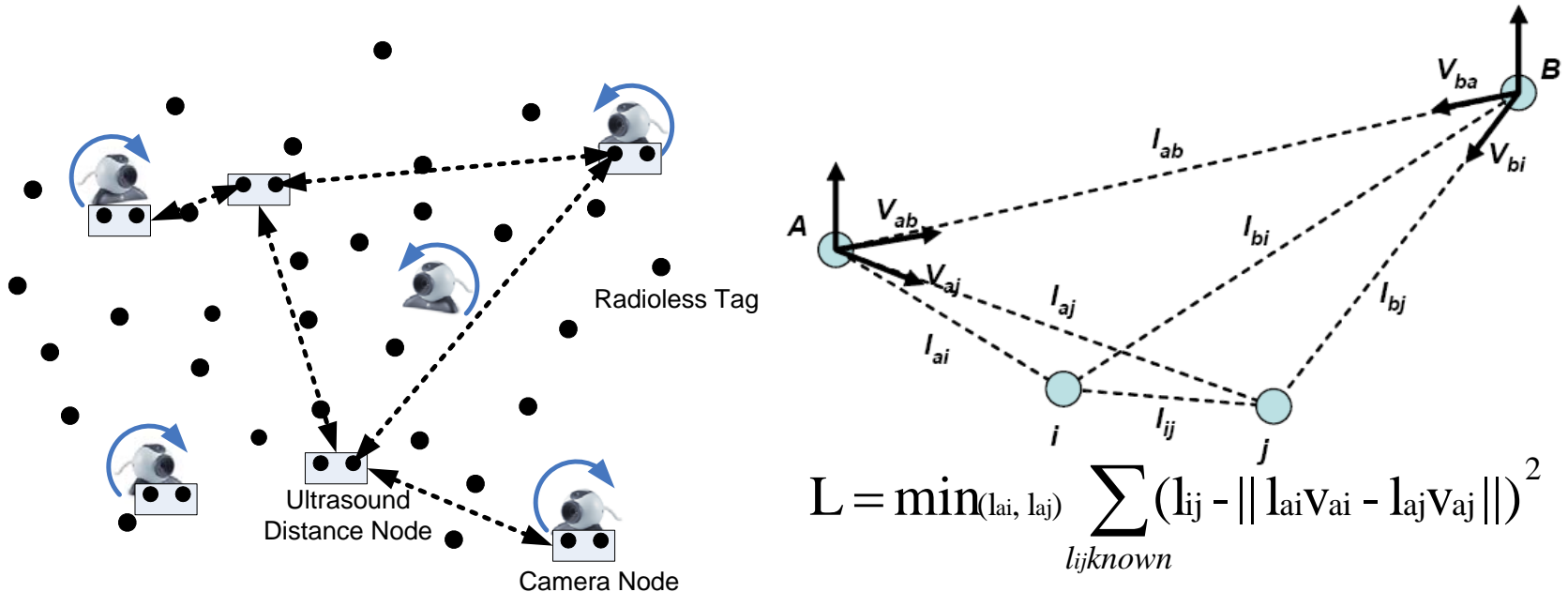


Pattern recognition with a single character

Network Design Implications

- **Extend address-event paradigm to the network**
 - Send selected spikes of events continuously to reduce the latency of a prediction
- **Adaptive fidelity in sensing and computation**
 - **First:** what & where
 - **Second:** image acquisition on demand
- **Minimize software maintenance at the leafs**
- **Drive**

Experiments with OV - Node Localization and Camera Calibration

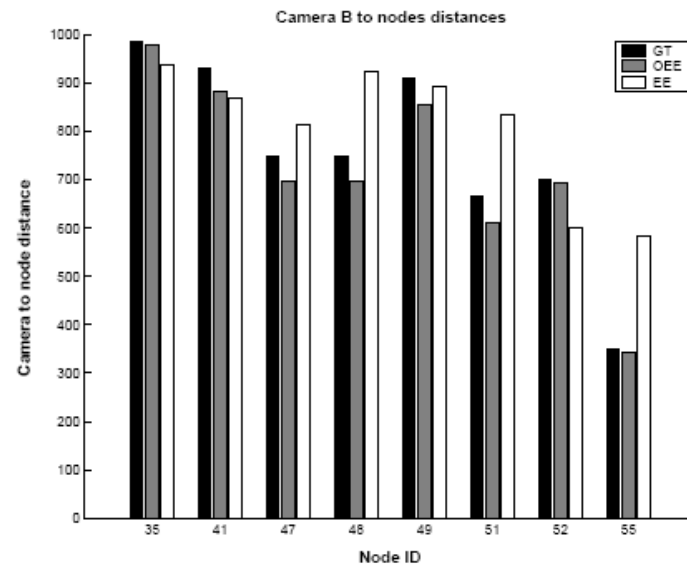
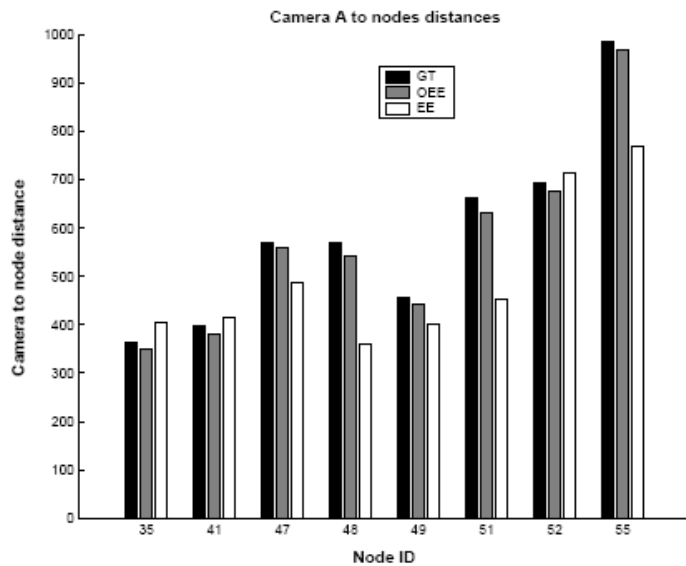
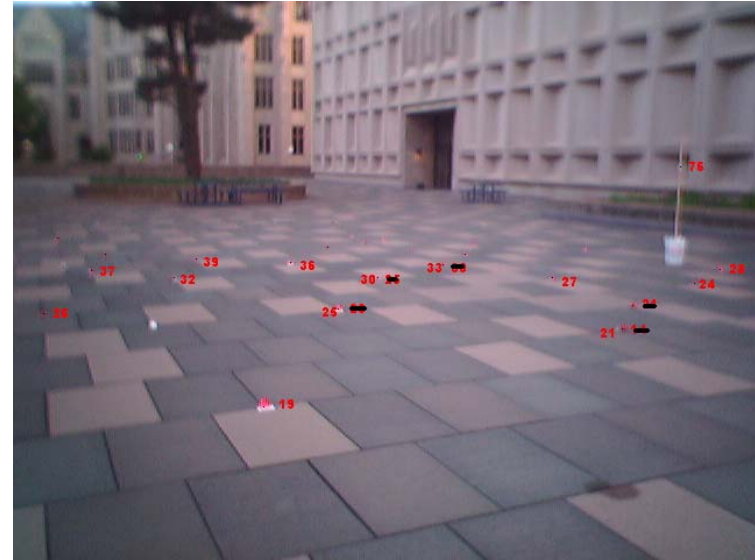
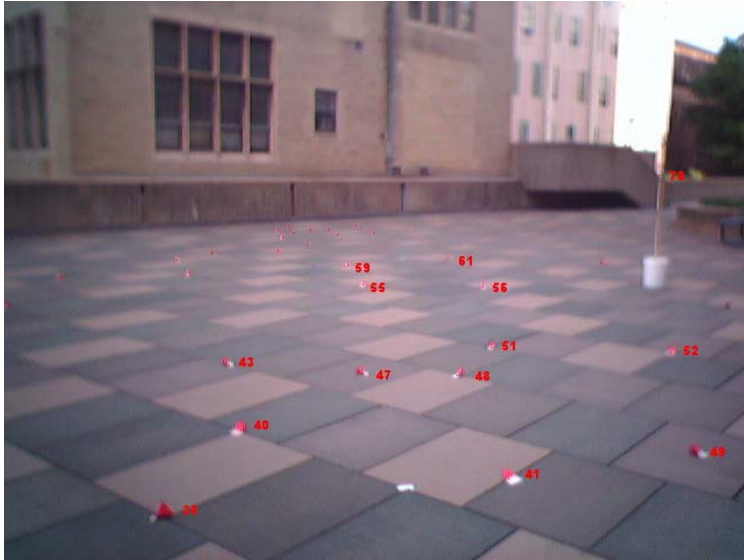


$$L = \min(l_{ai}, l_{aj}) \sum_{l_{ij} \text{ known}} (l_{ij} - \|l_{ai}V_{ai} - l_{aj}V_{aj}\|)^2$$

- Exploit information from only 2 camera views
- Setup cameras to understand depth w/o a distance measurement technology
- Light-weight computation using epipolar information
 - Measured Epipoles, Estimated Epipoles

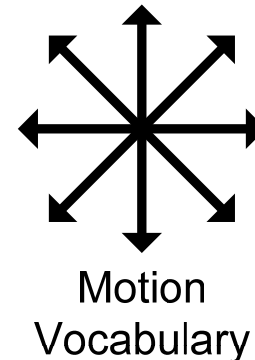
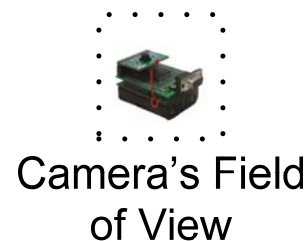
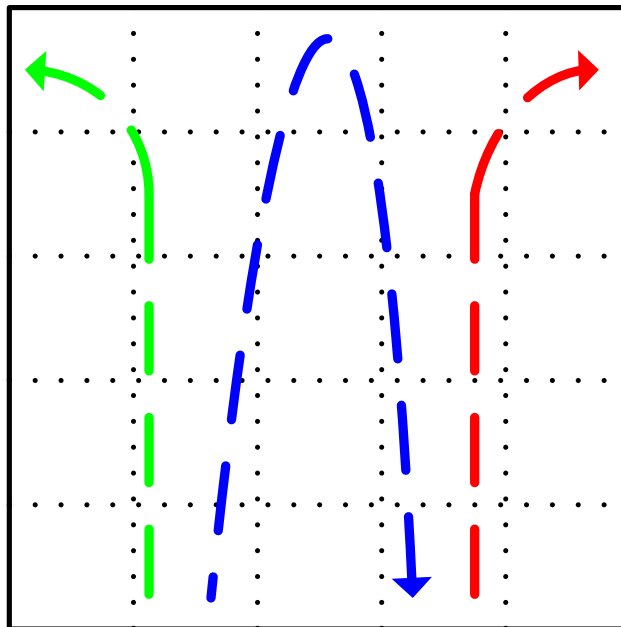
[D. Lymberopoulos, A. Barton-Sweeny and A. Savvides, "Sensor Localization and Camera Calibration Using Low Resolution Cameras"]

Results from Outdoor Deployment (Using OV+XYZ module)



Understanding Motion Behaviors in Physical Space

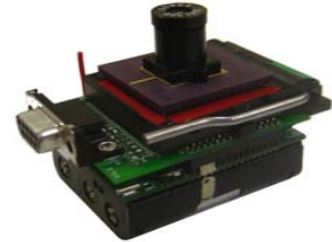
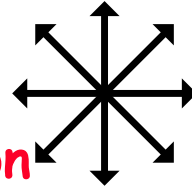
- Interested classifying tracking information into a more symbolic form
- **Goal:** Detect **global** motion patterns using **multiple locally** observed motion vectors
 - E.g left-turn, right-turn, u-turn



Understanding Patterns in Tracking Data

- **Level 0: Motion measurement**

Sensor Feature Extraction



- **Level 1: Detection of straight lines and turns - Phoneme**

- Use the angle θ between consequent motion vectors
- Choose the pattern p^* (straight line or turn) according to:

$$p^* = \max_i (p(\text{pattern}_i | \theta)) = \max_i (p(\theta | \text{pattern}_i) * p(\text{pattern}_i))$$

- **Level 2: Model complex patterns as a sequence of straight lines and turns - Words and Sentences**

- Left to right HMM without self-loops
- Every time that there is a change in the pattern p^* at level 1 a new state observation is made which value is the new p^* .
- Select the pattern Λ^* that satisfies:

$$\Lambda^* = \max_i (p(\Lambda_i | p_1^*, \dots, p_n^*)) = \max_i (p(p_1^*, \dots, p_n^* | \Lambda_i) * p(\Lambda_i))$$

Potential Applications Tactical Environments

- **Omni-directional custom camera arrays**
 - Minimize computation, maximize information
 - Hardware filters that detect humans in a scene
- **Sensory Grammars**
 - Parse sensor data to text & summaries inside the network
 - What, when & where assessment before full image recovery
- **Network reconfiguration and role assignment**
 - Write a global specification in sensory grammar
 - Automatically diffuse and reconfigure the **cognitive hierarchy** of the network

Conclusions

- **Filter information at the sensor**
 - Consider architectures that extract only the useful bits of information out of a scene
 - HW filters for detecting only humans
- **Platform implications**
 - Need platforms that can handle low cost imagers and acoustic at the same time (better architectures ARE possible)
 - PCB is the dominant cost. Cost will go down w/ single chip production - discussing SOC with OKI
- **Two main methods of reconfiguration**
 - SOS Operating System Dynamic Module
 - Sensory Grammar - (under development)
- **Successful Complex Systems: Make sensors SMALLER, CHEAPER, SIMPLER, POWER EFFICIENT AND MORE INTELLIGENT!!!**